

# Asymptotic Properties of Minimum $S$ -Divergence Estimator for Discrete Models \*

Abhik Ghosh  
 Indian Statistical Institute  
 203 B. T. Road, Kolkata 700 108, India  
*abhianik@gmail.com*

## Abstract

Robust inference based on the minimization of statistical divergences has proved to be a useful alternative to the classical techniques based on maximum likelihood and related methods. Recently Ghosh et al. (2013) proposed a general class of divergence measures, namely the  $S$ -Divergence Family and discussed its usefulness in robust parametric estimation through some numerical illustrations. In this present paper, we develop the asymptotic properties of the proposed minimum  $S$ -Divergence estimators under discrete models.

**Keywords:**  $S$ -Divergence Robustness Asymptotic Normality

## 1 Introduction

Density-based minimum distance methods provide attractive alternatives to likelihood based methods in parametric inference. Often these estimators combine strong robustness properties with full asymptotic efficiency. The estimators based on the family of power divergences (Cressie and Read, 1984) is one such example. The power divergence measure between two densities  $g$  and  $f$ , indexed by a parameter  $\lambda \in \mathbb{R}$ , is defined as

$$PD_\lambda(g, f) = \frac{1}{\lambda(\lambda + 1)} \int g[(g/f)^\lambda - 1].$$

For values of  $\lambda = 1, 0, -1/2, -1$  and  $-2$  the family generates the Pearson's chi-square (PCS), the likelihood disparity (LD), the Hellinger distance (HD), the Kullback-Leibler divergence (KLD) and the Neyman's chi-square (NCS) respectively. The family is a subclass of the larger family of  $\phi$ -divergences (Csiszár, 1963) or disparities. The minimum disparity estimator of  $\theta$  under the model  $\mathcal{F} = \{F_\theta : \theta \in \Theta \subseteq \mathbb{R}^p\}$  is the minimizer of the divergence between  $\hat{g}$  (a nonparametric estimate of the true density  $g$ ) and the model density  $f_\theta$ . All minimum distance estimators based on disparities have the same influence function as that of the maximum likelihood estimator (MLE) at the model and hence have the same asymptotic model efficiency.

The evaluation of a minimum distance estimator based on disparities requires kernel density estimation, and hence inherits all the complications of the latter method. Basu et al. (1998) developed a class of density-based divergence measures called the density power divergence (DPD) that produces robust parameter estimates but needs no nonparametric smoothing. The DPD measure between two densities  $g$  and  $f$  is

---

\*This is part of the Ph.D. research work of the author which is ongoing under Prof. Ayanendranath Basu at the Indian Statistical Institute

defined, depending on a nonnegative parameter  $\alpha$ , as

$$d_\alpha(g, f) = \int f^{1+\alpha} - \frac{1+\alpha}{\alpha} \int f^\alpha g + \frac{1}{\alpha} \int g^{1+\alpha}, \quad \text{for } \alpha > 0,$$

and

$$d_0(g, f) = \lim_{\alpha \rightarrow 0} d_\alpha(g, f) = \int g \log(g/f). \quad (1)$$

The parameter  $\alpha$  provides a smooth bridge between the likelihood disparity ( $\alpha = 0$ ) and the  $L_2$ -divergence ( $\alpha = 1$ ); it also controls the trade-off between the robustness and efficiency with larger  $\alpha$  being associated with greater robustness but reduced efficiency. Both the PD and DPD families provide outlier down weighting using powers of model densities.

Combining the concepts of the power divergence and the density power divergence, Ghosh et al. (2013) developed a two parameter family of density-based divergences, named as “ $S$ -Divergence”, that connects the whole of the Cressie-Read family of power divergence smoothly to the  $L_2$ -divergence at the other end. This family contains both the PD and DPD families as special cases. Through various numerical examples, they illustrate that the minimum divergence estimators based on the  $S$ -Divergence are also extremely robust and are also competitive in terms efficiency for most of the members of this family.

In this present article, we will develop the theoretical properties of the minimum  $S$ -Divergence estimators. For simplicity, here we consider only the set up for the discrete model so that the true data generating probability mass functions can be estimated non-parametrically by just the relative frequencies of the observed sample — we do not need to consider any nonparametric smoothing. We will prove the consistency and asymptotic normality of the minimum  $S$ -Divergence estimators. We will introduce the  $S$ -divergence and the minimum  $S$ -divergence estimator in Section 2 and 3 respectively. Then Section 4 will contain the asymptotic properties of the minimum  $S$ -Divergence estimators. Finally we conclude the paper by examining a real data example in Section 5 and an overall conclusion in Section 6. Throughout the rest of the paper, we will use the term “density” for the probability mass functions also.

## 2 The $S$ -Divergence Family

It is well-known that the estimating equation for the minimum density power divergence represents an interesting density power down-weighting, and hence robustification of the usual likelihood score equation (Basu et al., 1998). The usual estimating equations for the MLE can be recovered from that estimating equation by the choice  $\alpha = 0$ . Within the given range of  $0 \leq \alpha \leq 1$ ,  $\alpha = 1$  will lead to the maximum down-weighting for the score functions of the surprising observations corresponding to the  $L_2$  divergence ; on the other extreme, the score functions will be subjected to no down-weighting at all for  $\alpha = 0$  corresponding to the Kullback-Leibler divergence (Kullback and Leibler, 1951). Intermediate values of  $\alpha$  provide a smooth bridge between these two estimating equations, and the degree of down weighting increases with increasing  $\alpha$ .

Noting that the Kullback-Leibler divergence is a particular case of Cressie-Read family of power divergence corresponding to  $\lambda = 0$ , we see that the density power divergence gives us a smooth bridge between one particular member of the Cressie-Read family and the  $L_2$  divergence with increasing robustness. Ghosh et al. (2013) constructed a family of divergences which connect, in a similar fashion, other members of the PD family with the  $L_2$  distance. That larger super-family, named as the  $S$ -Divergence Family, is defined as

$$\begin{aligned}
S_{(\alpha,\lambda)}(g, f) &= \frac{1}{1 + \lambda(1 - \alpha)} \int [(f^{1+\alpha} - g^{1+\alpha}) \\
&\quad - \frac{(1 + \alpha)}{(\alpha - \lambda(1 - \alpha))} g^{1+\lambda(1-\alpha)} (f^{\alpha-\lambda(1-\alpha)} - g^{\alpha-\lambda(1-\alpha)})] \\
&= \frac{1}{A} \int f^{1+\alpha} - \frac{1 + \alpha}{AB} \int f^B g^A + \frac{1}{B} \int g^{1+\alpha},
\end{aligned} \tag{2}$$

where  $A = 1 + \lambda(1 - \alpha)$  and  $B = \alpha - \lambda(1 - \alpha)$ .

Note that,  $A + B = 1 + \alpha$ . Also the above form of divergence family is defined for those  $\alpha$  and  $\lambda$  for which  $A \neq 0$  and  $B \neq 0$ . If  $A = 0$  then the corresponding divergence measure is defined as the continuous limit of (2) as  $A \rightarrow 0$  and is given by

$$\begin{aligned}
S_{(\alpha,\lambda:A=0)}(g, f) &= \lim_{A \rightarrow 0} S_{(\alpha,\lambda)}(g, f) \\
&= \int f^{1+\alpha} \log\left(\frac{f}{g}\right) - \int \frac{(f^{1+\alpha} - g^{1+\alpha})}{1 + \alpha}.
\end{aligned} \tag{3}$$

Similarly, if  $B = 0$  then the divergence measure is defined to be

$$\begin{aligned}
S_{(\alpha,\lambda:B=0)}(g, f) &= \lim_{B \rightarrow 0} S_{(\alpha,\lambda)}(g, f) \\
&= \int g^{1+\alpha} \log\left(\frac{g}{f}\right) - \int \frac{(g^{1+\alpha} - f^{1+\alpha})}{1 + \alpha}.
\end{aligned} \tag{4}$$

Note that for  $\alpha = 0$ , the class of  $S$ -divergences reduces to the PD family with parameter  $\lambda$ ; for  $\alpha = 1$ ,  $S_{1,\lambda}$  equals the  $L_2$  divergence irrespective of the value of  $\lambda$ . On the other hand,  $\lambda = 0$  generates the DPD family as a function of  $\alpha$ . In Ghosh et al. (2013), it was shown that the above  $S$ -divergence family defined in (2), (3) and (4) indeed represent a family of genuine statistical divergence measures in the sense that  $S_{(\alpha,\lambda)}(g, f) \geq 0$  for densities  $g, f$  and all  $\alpha \geq 0$ ,  $\lambda \in \mathbb{R}$ , and  $S_{(\alpha,\lambda)}(g, f)$  is equal to zero if and only if  $g = f$  identically.

### 3 The Minimum $S$ -Divergence Estimators

Let us now consider the discrete set-up for parametric estimation. Let  $X_1, \dots, X_n$  denotes  $n$  independent and identically distributed observations from the true distribution  $G$  having a probability density function  $g$  with respect to some counting measure. Without loss of generality, we will assume that the support of  $g$  is  $\chi = \{0, 1, 2, \dots\}$ . Let us denote the relative frequency at  $x$  obtained from data by  $r_n(x) = \frac{1}{n} \sum_{i=1}^n \chi(X_i = x)$ . We model the true data generating distribution  $G$  by the parametric model family  $\mathcal{F} = \{F_\theta : \theta \in \Theta \subseteq \mathbb{R}^p\}$ . We will assume that both  $G$  and  $\mathcal{F}$  belong to  $\mathcal{G}$ , the (convex) class of all distributions having densities with respect to the counting measure (or the appropriate dominating measure in other cases). We are interested in the estimation of the parameter  $\theta$ .

Note that, the minimum  $S$ -divergence estimator has to be obtained by minimizing the  $S$ -divergence measure between the data and the model distribution. However, in the discrete set-up, both the data-generating true distribution and the model distribution are characterized by the probability vectors  $\mathbf{r}_n = (r_n(0), r_n(1), \dots)^T$  and  $\mathbf{f}_\theta = (f_\theta(0), f_\theta(1), \dots)^T$  respectively. Thus in this case, the minimum  $S$ -divergence estimator of  $\theta$  can be obtained by just minimizing  $S_{(\alpha,\lambda)}(\mathbf{r}_n, \mathbf{f}_\theta)$ , the  $S$ -divergence measure between  $\mathbf{r}_n$  and

$\mathbf{f}_\theta$ , with respect to  $\theta$ . The estimating equation is then given by

$$\sum_{x=0}^{\infty} f_\theta^{1+\alpha}(x) u_\theta(x) - \sum_{x=0}^{\infty} f_\theta^B(x) r_n^A(x) u_\theta(x) = 0, \quad (5)$$

$$\text{or, } \sum_{x=0}^{\infty} K(\delta(x)) f_\theta^{1+\alpha}(x) u_\theta(x) = 0, \quad (6)$$

where  $\delta(x) = \delta_n(x) = \frac{r_n(x)}{f_\theta(x)} - 1$ ,  $K(\delta) = \frac{(\delta+1)^A - 1}{A}$  and  $u_\theta(x) = \nabla \ln f_\theta(x)$  is the likelihood score function. Note that,  $\nabla$  represents the derivative with respect to  $\theta$  and we will denote its  $i^{\text{th}}$  component by  $\nabla_i$ .

## 4 Asymptotic properties of the Minimum $S$ -Divergence Estimators

Now we will derive the asymptotic properties of the minimum  $S$ -divergence estimator under the discrete set-up as mentioned above. Note that, in order to obtain the minimum  $S$ -divergence estimator under discrete set-up, we need to minimize  $S_{(\alpha, \lambda)}(\mathbf{r}_n, \mathbf{f}_\theta)$  over  $\theta$  which is equivalent to minimizing  $H_n(\theta)$  with respect to  $\theta$  where

$$H_n(\theta) = \frac{1}{1+\alpha} \left[ \frac{1}{A} \sum_x f_\theta^{1+\alpha}(x) - \frac{1+\alpha}{AB} \sum_x f_\theta^B(x) r_n^A(x) \right]. \quad (7)$$

Now,

$$\begin{aligned} \nabla H_n(\theta) &= \frac{1}{A} \left[ \sum_x f_\theta^{1+\alpha}(x) u_\theta(x) - \sum_x f_\theta^B(x) u_\theta(x) r_n^A(x) \right] \\ &= -\frac{1}{A} \sum_x K(\delta_n(x)) f_\theta^{1+\alpha}(x) u_\theta(x), \end{aligned} \quad (8)$$

where  $\delta_n(x) = \frac{r_n(x)}{f_\theta(x)} - 1$ . Thus the estimating equation is exactly the same as given in Equation (6). Let  $\theta^g$  denotes the “best fitting parameter” under the true density  $g$ , obtained by minimizing  $S_{(\alpha, \lambda)}(g, \mathbf{f}_\theta)$  over the parameter space  $\theta \in \Theta$ . Define

$$\begin{aligned} J_g &= J_\alpha(g) = E_g [u_{\theta^g}(X) u_{\theta^g}^T(X) K'(\delta_g^g(X)) f_{\theta^g}^\alpha(X)] \\ &\quad - \sum_x K(\delta_g^g(x)) \nabla^2 f_{\theta^g}(x), \end{aligned} \quad (9)$$

$$V_g = V_\alpha(g) = V_g [K'(\delta_g^g(X)) f_{\theta^g}^\alpha(X) u_{\theta^g}(X)], \quad (10)$$

where  $\delta_g(x) = \frac{g(x)}{f_{\theta^g}(x)} - 1$ ,  $\delta_g^g(x) = \frac{g(x)}{f_{\theta^g}(x)} - 1$ ,  $K'(\cdot)$  is the derivative of  $K(\cdot)$  with respect to its argument and  $\nabla^2$  represent the second order derivative with respect to  $\theta$ . We will prove the asymptotic properties of the minimum  $S$ -divergence estimator under the following assumptions:

(SA1) The model family  $\mathcal{F}$  is identifiable, i.e., for any two  $F_{\theta_1}$  and  $F_{\theta_2}$  in the model family  $\mathcal{F}$ ,

$$F_{\theta_1} = F_{\theta_2} \Rightarrow \theta_1 = \theta_2.$$

(SA2) The probability density function  $f_\theta$  of the model distribution have common support so that the set  $\chi = \{x : f_\theta(x) > 0\}$  is independent of  $\theta$ . Also the true distribution  $g$  is compatible with the model family.

- (SA3) There exists an open subset  $\omega \subset \Theta$  for which the best fitting parameter  $\theta^g$  is an interior point and for almost all  $x$ , the density  $f_\theta(x)$  admits all the third derivatives of the type  $\nabla_{jkl} f_\theta(x) \forall \theta \in \omega$ . Here,  $\nabla_{jkl}$  denotes the  $(j, k, l)^{\text{th}}$  element of  $\nabla^3$ , the third order derivative with respect to  $\theta$
- (SA4) The matrix  $\frac{1+\alpha}{A} J_g$  is positive definite.
- (SA5) The quantities  $\sum_x g^{1/2}(x) f_\theta^\alpha(x) |u_{j\theta}(x)|$ ,  $\sum_x g^{1/2}(x) f_\theta^\alpha(x) |u_{j\theta}(x)| |u_{k\theta}(x)|$  and  $\sum_x g^{1/2}(x) f_\theta^\alpha(x) |u_{jk\theta}(x)|$  are bounded  $\forall j, k$  and  $\forall \theta \in \omega$ .  
Here,  $u_{j\theta}(x)$  denotes the  $j^{\text{th}}$  element of  $u_\theta(x)$  and  $u_{jk\theta}(x)$  denotes the  $(j, k)^{\text{th}}$  element of  $\nabla^2 \ln f_\theta(x)$ .
- (SA6) For almost all  $x$ , there exists functions  $M_{jkl}(x)$ ,  $M_{jk,l}(x)$ ,  $M_{j,k,l}(x)$  that dominate, in absolute value,  $f_\theta^\alpha(x) u_{jkl\theta}(x)$ ,  $f_\theta^\alpha(x) u_{jk\theta}(x) u_{l\theta}(x)$  and  $f_\theta^\alpha(x) u_{j\theta}(x) u_{k\theta}(x) u_{l\theta}(x)$  respectively  $\forall j, k, l$  and that are uniformly bounded in expectation with respect to  $g$  and  $f_\theta \forall \theta \in \omega$ .  
Here,  $u_{jkl\theta}(x)$  denotes the  $(j, k, l)^{\text{th}}$  element of  $\nabla^3 \ln f_\theta(x)$ .
- (SA7) The function  $\left(\frac{g(x)}{f_\theta(x)}\right)^{A-1}$  is uniformly bounded (by, say,  $C$ )  $\forall \theta \in \omega$ .

To prove the consistency and asymptotic normality of the minimum  $S$ -divergence estimator, we will, now on, assume that the above 7 conditions hold. We will first consider some Lemmas.

**Lemma 4.1** Define  $\eta_n(x) = \sqrt{n}(\sqrt{\delta_n(x)} - \sqrt{\delta_g(x)})^2$ . For any  $k \in [0, 2]$  and any  $x \in \chi$ , we have

$$1. E_g[\eta_n(x)^k] \leq n^{\frac{k}{2}} E_g[|\delta_n(x) - \delta_g(x)|^k] \leq \left[ \frac{g(x)(1-g(x))}{f_\theta^2(x)} \right]^{\frac{k}{2}}.$$

$$2. E_g[|\delta_n(x) - \delta_g(x)|] \leq \frac{2g(x)(1-g(x))}{f_\theta(x)}.$$

**Proof :** For  $a, b \geq 0$ , we have the inequality  $(\sqrt{a} - \sqrt{b})^2 \leq |a - b|$ . So we get

$$\begin{aligned} E_g[\eta_n(x)^k] &= n^{\frac{k}{2}} E_g[(\sqrt{\delta_n(x)} - \sqrt{\delta_g(x)})^2]^k \\ &\leq n^{\frac{k}{2}} E_g[|\delta_n(x) - \delta_g(x)|^k]. \end{aligned}$$

For the next part see that, under  $g$ ,  $nr_n(x) \sim \text{Binomial}(n, g(x)) \forall x$ . Now, for any  $k \in [0, 2]$ , we get by the Lyapounov's inequality that

$$\begin{aligned} E_g[|\delta_n(x) - \delta_g(x)|^k] &\leq [E_g(\delta_n(x) - \delta_g(x))^2]^{\frac{k}{2}} \\ &= \frac{1}{f_\theta^k(x)} [E_g(r_n(x) - g(x))^2]^{\frac{k}{2}} \\ &= \frac{1}{f_\theta^k(x)} \left[ \frac{g(x)(1-g(x))}{n} \right]^{\frac{k}{2}}. \end{aligned}$$

For the second part, note that

$$\begin{aligned} E_g[|\delta_n(x) - \delta_g(x)|] &= \frac{1}{f_\theta^k(x)} [E_g|r_n(x) - g(x)|]^{\frac{k}{2}} \\ &\leq \frac{2g(x)(1-g(x))}{f_\theta(x)}, \end{aligned}$$

where the last inequality follows from the result about the mean-deviation of a Binomial random variable.  $\square$

**Lemma 4.2**  $E_g[\eta_n(x)^k] \rightarrow 0$ , as  $n \rightarrow \infty$ , for  $k \in [0, 2)$  and  $x \in \chi$ .

**Proof :** This follows from Theorem 4.5.2 of Chung (1974) by noting that  $n^{1/4}(r_n^{1/2}(x) - g^{1/2}(x)) \rightarrow 0$  with probability one for each  $x \in \chi$  and by the Lemma 4.1(1),  $\sup_n E_g[\eta_n^k(x)]$  is bounded.  $\square$

Let us now define,

$$\begin{aligned} a_n(x) &= K(\delta_n(x)) - K(\delta_g(x)), \\ b_n(x) &= (\delta_n(x) - \delta_g(x))K'(\delta_g(x)), \\ \text{and } \tau_n(x) &= \sqrt{n}|a_n(x) - b_n(x)|. \end{aligned}$$

We will need the limiting distributions of

$$S_{1n} = \sqrt{n} \sum_x a_n(x) f_\theta^{1+\alpha}(x) u_\theta(x) \text{ and } S_{2n} = \sqrt{n} \sum_x b_n(x) f_\theta^{1+\alpha}(x) u_\theta(x).$$

Next two Lemmas will help us to derive those distributions.

**Lemma 4.3** Assume condition (SA5). Then,

$$E_g|S_{1n} - S_{2n}| \rightarrow 0, \quad \text{as } n \rightarrow \infty,$$

and hence

$$S_{1n} - S_{2n} \xrightarrow{\mathcal{P}} 0, \quad \text{as } n \rightarrow \infty.$$

**Proof :** By Lemma 2.15 of Basu et al. (2011) [or, Lindsay (1994), Lemma 25], there exists some positive constant  $\beta$  such that

$$\tau_n(x) \leq \beta \sqrt{n} \left( \sqrt{\delta_n(x)} - \sqrt{\delta_g(x)} \right)^2 = \beta \eta_n(x).$$

Also, by Lemma 4.1,  $E_g[\tau_n(x)] \leq \beta \frac{g^{1/2}(x)}{f_\theta(x)}$ .

And by Lemma 4.2,  $E_g[\tau_n(x)] = \beta E_g[\eta_n(x)] \rightarrow 0$  as  $n \rightarrow \infty$ . Thus we get,

$$\begin{aligned} E_g|S_{1n} - S_{2n}| &\leq \sum_x E_g[\tau_n(x)] f_\theta^{1+\alpha}(x) |u_\theta(x)| \\ &\leq \beta \sum_x g^{1/2}(x) f_\theta^\alpha(x) |u_\theta(x)| \\ &< \infty \quad (\text{by assumption SA5}). \end{aligned}$$

So, by Dominated Convergence Theorem (DCT),  $E_g|S_{1n} - S_{2n}| \rightarrow 0$  as  $n \rightarrow \infty$ .

Hence, by Markov inequality,  $S_{1n} - S_{2n} \xrightarrow{\mathcal{P}} 0$  as  $n \rightarrow \infty$ .  $\square$

**Lemma 4.4** Suppose  $V_g$  is finite. Then under  $g$ ,

$$S_{1n} \xrightarrow{\mathcal{D}} N(0, V_g).$$

**Proof :** Note that, by the previous Lemma 4.3, the asymptotic distribution of  $S_{1n}$  and  $S_{2n}$  are the same. Now, we have

$$\begin{aligned}
S_{2n} &= \sqrt{n} \sum_x (\delta_n(x) - \delta_g(x)) K'(\delta_g(x)) f_\theta^{1+\alpha}(x) u_\theta(x) \\
&= \sqrt{n} \sum_x (r_n(x) - g(x)) K'(\delta_g(x)) f_\theta^\alpha(x) u_\theta(x) \\
&= \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n [K'(\delta_g(X_i)) f_\theta^\alpha(X_i) u_\theta(X_i) - E_g\{K'(\delta_g(X)) f_\theta^\alpha(X) u_\theta(X)\}] \right) \\
&\Rightarrow Z \sim N(0, V_g) \quad [\text{by Central Limit Theorem (CLT)}].
\end{aligned}$$

This completes the proof.  $\square$

We will now consider the main theorem of this section about the consistency and asymptotic normality of the minimum  $S$ -divergence estimator.

**Theorem 4.5** *Under Assumptions (SA1)-(SA7), there exists a consistent sequence  $\hat{\theta}_n$  of roots to the minimum  $S$ -divergence estimating equation (6).*

*Also, the asymptotic distribution of  $\sqrt{n}(\hat{\theta}_n - \theta^g)$  is  $p$ -dimensional normal with mean 0 and variance  $J_g^{-1} V_g J_g^{-1}$ .*

**Proof of consistency:** Consider the behavior of  $S_{(\alpha, \lambda)}(\mathbf{r}_n, \mathbf{f}_\theta)$  on a sphere  $Q_a$  which has radius  $a$  and center at  $\theta^g$ . We will show, for sufficiently small  $a$ , the probability tends to one that

$$S_{(\alpha, \lambda)}(\mathbf{r}_n, \mathbf{f}_\theta) > S_{(\alpha, \lambda)}(\mathbf{r}_n, \mathbf{f}_{\theta^g}) \quad \forall \theta \quad \text{on the surface of } Q_a,$$

so that the  $S$ -divergence has a local minimum with respect to  $\theta$  in the interior of  $Q_a$ . At a local minimum, the estimating equations must be satisfied. Therefore, for any  $a > 0$  sufficiently small, the minimum  $S$ -divergence estimating equation have a solution  $\theta_n$  within  $Q_a$  with probability tending to one as  $n \rightarrow \infty$ .

Now taking Taylor series expansion of  $S_{(\alpha, \lambda)}(\mathbf{r}_n, \mathbf{f}_\theta)$  about  $\theta = \theta^g$ , we get

$$\begin{aligned}
&S_{(\alpha, \lambda)}(\mathbf{r}_n, \mathbf{f}_\theta) - S_{(\alpha, \lambda)}(\mathbf{r}_n, \mathbf{f}_{\theta^g}) \\
&= - \sum_j (\theta_j - \theta_j^g) \nabla_j S_{(\alpha, \lambda)}(\mathbf{r}_n, \mathbf{f}_\theta)|_{\theta=\theta^g} \\
&\quad - \frac{1}{2} \sum_{j,k} (\theta_j - \theta_j^g)(\theta_k - \theta_k^g) \nabla_{jk} S_{(\alpha, \lambda)}(\mathbf{r}_n, \mathbf{f}_\theta)|_{\theta=\theta^g} \\
&\quad - \frac{1}{6} \sum_{j,k,l} (\theta_j - \theta_j^g)(\theta_k - \theta_k^g)(\theta_l - \theta_l^g) \nabla_{jkl} S_{(\alpha, \lambda)}(\mathbf{r}_n, \mathbf{f}_\theta)|_{\theta=\theta^*} \\
&= S_1 + S_2 + S_3, \quad (\text{say})
\end{aligned}$$

where  $\theta^*$  lies between  $\theta^g$  and  $\theta$ . We will now consider each terms one-by-one.

For the Linear term  $S_1$ , we consider

$$\nabla_j S_{(\alpha, \lambda)}(\mathbf{r}_n, \mathbf{f}_\theta)|_{\theta=\theta^g} = - \frac{(1+\alpha)}{A} \sum_x K(\delta_n^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x), \quad (11)$$

where  $\delta_n^g(x)$  is the  $\delta_n(x)$  evaluated at  $\theta = \theta^g$ . We will now show that

$$\sum_x K(\delta_n^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x) \xrightarrow{P} \sum_x K(\delta_g^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x), \quad (12)$$

as  $n \rightarrow \infty$  and note that the right hand side of above is zero by definition of the Minimum  $S$ -divergence estimator. Note that by assumption (SA7) and the fact that  $r_n(x) \rightarrow g(x)$ , almost surely (a.s.) by Strong Law of Large Number (SLLN), it follows that

$$|K'(\delta)| = |A||\delta|^{A-1} < 2|A|C = C_1, \quad (\text{say}) \quad (13)$$

for any  $\delta$  in between  $\delta_n^g(x)$  and  $\delta_g^g(x)$ , uniformly in  $x$ . So, by using the one-term Taylor series expansion,

$$\begin{aligned} & \left| \sum_x K(\delta_n^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x) - \sum_x K(\delta_g^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x) \right| \\ & \leq C_1 \sum_x |\delta_n^g(x) - \delta_g^g(x)| f_{\theta^g}^{1+\alpha}(x) |u_{j\theta^g}(x)|. \end{aligned}$$

However, by Lemma 4.1(1),

$$E[|\delta_n^g(x) - \delta_g^g(x)|] \leq \frac{(g(x)(1-g(x)))^{1/2}}{f_{\theta^g}(x)\sqrt{n}} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (14)$$

and, by Lemma 4.1(2), we have

$$\begin{aligned} & E[C_1 \sum_x |\delta_n^g(x) - \delta_g^g(x)| f_{\theta^g}^{1+\alpha}(x) |u_{j\theta^g}(x)|] \\ & \leq 2C_1 \sum_x g^{1/2}(x) f_{\theta^g}^\alpha(x) |u_{j\theta^g}(x)| < \infty. \end{aligned} \quad (15)$$

[by assumption (A5)]

Hence, by DCT, we get,

$$E\left[\left| \sum_x K(\delta_n^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x) - \sum_x K(\delta_g^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x) \right|\right] \rightarrow 0, \quad (16)$$

as  $n \rightarrow \infty$ , so that by Markov inequality we have the desired claim. Therefore, we have

$$\nabla_j S_{(\alpha, \lambda)}(\mathbf{r}_n, \mathbf{f}_\theta)|_{\theta=\theta^g} \xrightarrow{\mathcal{P}} 0. \quad (17)$$

Thus, with probability tending to one,  $|S_1| < pa^3$ , where  $p$  is the dimension of  $\theta$  and  $a$  is the radius of  $Q_a$ .

Next we consider the quadratic term  $S_2$ . We have,

$$\begin{aligned} & \nabla_{jk} S_{(\alpha, \lambda)}(r_n, f_\theta)|_{\theta=\theta^g} \\ & = \nabla_k \left( -\frac{(1+\alpha)}{A} \sum_x K(\delta_n(x)) f_\theta^{1+\alpha}(x) u_{j\theta}(x) \right)|_{\theta=\theta^g} \\ & = -\frac{(1+\alpha)}{A} \left[ -\sum_x K'(\delta_n^g(x)) \delta_n^g(x) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x) u_{k\theta^g}(x) \right. \\ & \quad \left. + \sum_x K(\delta_n^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{jk\theta^g}(x) \right. \\ & \quad \left. - \sum_x K(\delta_n^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x) u_{k\theta^g}(x) \right]. \end{aligned} \quad (18)$$



We will now show that

$$\begin{aligned} & - \sum_x K'(\delta_n^g(x)) \delta_n^g(x) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x) u_{k\theta^g}(x) \\ & \xrightarrow{\mathcal{P}} - \sum_x K'(\delta_g^g(x)) \delta_g^g(x) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x) u_{k\theta^g}(x). \end{aligned} \quad (19)$$

For note that as in (13), we have

$$|K''(\delta)\delta| = |A(A-1)| |\delta|^{(A-1)} < C_2, \quad (\text{say}) \quad (20)$$

for every  $\delta$  lying in between  $\delta_n^g(x)$  and  $\delta_g^g(x)$ , uniformly in  $x$ . So, by using the one-term Taylor series expansion,

$$\begin{aligned} |K'(\delta_n^g) \delta_n^g - K'(\delta_g^g) \delta_g^g| & \leq |\delta_n^g - \delta_g^g| |K''(\delta_n^*) \delta_n^* + K'(\delta_n^*)| \\ & \leq |\delta_n^g - \delta_g^g| (C_2 + C_1). \end{aligned}$$

Thus, we get

$$\begin{aligned} & \left| \sum_x K'(\delta_n^g(x)) \delta_n^g(x) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x) u_{k\theta^g}(x) \right. \\ & \quad \left. - \sum_x K'(\delta_g^g(x)) \delta_g^g(x) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x) u_{k\theta^g}(x) \right| \\ & \leq (C_1 + C_2) \sum_x |\delta_n^g - \delta_g^g| f_{\theta^g}^{1+\alpha}(x) |u_{j\theta^g}(x) u_{k\theta^g}(x)|. \end{aligned}$$

Since by assumption (SA5), we have  $\sum_x g^{1/2}(x) f_{\theta^g}^{1+\alpha}(x) |u_{j\theta^g}(x) u_{k\theta^g}(x)| < \infty$ , the desired result (19) follows by the similar proof for proving (12) above. Similarly we also get that

$$\sum_x K(\delta_n^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{jk\theta^g}(x) \xrightarrow{\mathcal{P}} \sum_x K(\delta_g^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{jk\theta^g}(x), \quad (21)$$

$$\sum_x K(\delta_n^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x) u_{k\theta^g}(x) \xrightarrow{\mathcal{P}} \sum_x K(\delta_g^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{j\theta^g}(x) u_{k\theta^g}(x). \quad (22)$$

Thus, combining (19), (21) and (22), we get that,

$$\nabla_k \left( \sum_x K(\delta_n(x)) f_{\theta}^{1+\alpha}(x) u_{j\theta}(x) |_{\theta=\theta^g} \right) \xrightarrow{\mathcal{P}} -J_g^{j,k}. \quad (23)$$

But we have,

$$\begin{aligned} 2S_2 &= \frac{1+\alpha}{A} \sum_{j,k} \left\{ \nabla_k \left( \sum_x K(\delta_n(x)) f_{\theta}^{1+\alpha}(x) u_{j\theta}(x) |_{\theta=\theta^g} \right) - (-J_g^{j,k}) \right\} \\ & \quad \times (\theta_j - \theta_j^g)(\theta_k - \theta_k^g) \\ & \quad + \sum_{j,k} \left\{ - \left( \frac{1+\alpha}{A} J_g^{j,k} \right) (\theta_j - \theta_j^g)(\theta_k - \theta_k^g) \right\}. \end{aligned} \quad (24)$$

Now the absolute value of the first term in above (24) is  $< p^2 a^3$  with probability tending to one. And, the second term in (24) is a negative definite quadratic form in the variables  $(\theta_j - \theta_j^g)$ . Letting  $\lambda_1$  be the largest eigenvalue of  $\frac{(1+\alpha)}{A} J_g$ , the quadratic form is  $< \lambda_1 a^2$ . Combining the two terms, we see that there exists  $c > 0$  and  $a_0 > 0$  such that for  $a < a_0$ , we have  $S_2 < -ca^2$  with probability tending to one.

Finally, considering the cubic term  $S_3$ , we have

$$\begin{aligned} \nabla_{jkl} S_{(\alpha, \lambda)}(r_n, f_\theta)|_{\theta=\theta^*} &= \nabla_{kl} \left( -\frac{(1+\alpha)}{A} \sum_x K(\delta_n(x)) f_\theta^{1+\alpha}(x) u_{j\theta}(x) \right)|_{\theta=\theta^*} \\ &= -\frac{(1+\alpha)}{A} \nabla_l \left( -\sum_x K'(\delta_n(x)) \delta_n(x) f_\theta^{1+\alpha}(x) u_{j\theta}(x) u_{k\theta}(x) \right. \\ &\quad \left. + \sum_x K(\delta_n(x)) f_\theta^\alpha(x) \nabla_{jk} f_\theta(x) \right)|_{\theta=\theta^*}, \end{aligned} \quad (25)$$

or,

$$\begin{aligned} &-\frac{A}{1+\alpha} \nabla_{jkl} S_{(\alpha, \lambda)}(r_n, f_\theta)|_{\theta=\theta^*} \\ &= \sum_x K''(\delta_n^*(x)) \delta_n^*(x)^2 f_{\theta^*}^{1+\alpha}(x) u_{j\theta^*}(x) u_{k\theta^*}(x) u_{l\theta^*}(x) \\ &\quad - \sum_x K'(\delta_n^*(x)) \delta_n^*(x) f_{\theta^*}^{1+\alpha}(x) u_{j\theta^*}(x) u_{k\theta^*}(x) u_{l\theta^*}(x) \\ &\quad - \sum_x K'(\delta_n^*(x)) \delta_n^*(x) f_{\theta^*}^{1+\alpha}(x) u_{j\theta^*}(x) u_{kl\theta^*}(x) \\ &\quad - \sum_x K'(\delta_n^*(x)) \delta_n^*(x) f_{\theta^*}^{1+\alpha}(x) u_{k\theta^*}(x) u_{jl\theta^*}(x) \\ &\quad - \sum_x K'(\delta_n^*(x)) \delta_n^*(x) f_{\theta^*}^{1+\alpha}(x) \frac{\nabla_{jk} f_{\theta^*}(x)}{f_{\theta^*}(x)} u_{l\theta^*}(x) \\ &\quad + \sum_x K(\delta_n^*(x)) f_{\theta^*}^{1+\alpha}(x) \frac{\nabla_{jk} f_{\theta^*}(x)}{f_{\theta^*}(x)} \\ &\quad + \sum_x K(\delta_n^*(x)) f_{\theta^*}^{1+\alpha}(x) \frac{\nabla_{jk} f_{\theta^*}(x)}{f_{\theta^*}(x)} u_{l\theta^*}(x), \end{aligned} \quad (26)$$

where  $\delta_n^*(x) = \frac{r_n(x)}{f_{\theta^*}(x)}$ . We will now show that all the terms in the RHS of above (26) are bounded. Let us name the terms by (i) to (vii) respectively. For the first term (i), we use (20) to get

$$\begin{aligned} &|\sum_x K''(\delta_n^*(x)) \delta_n^*(x)^2 f_{\theta^*}^{1+\alpha}(x) u_{j\theta^*}(x) u_{k\theta^*}(x) u_{l\theta^*}(x)| \\ &\leq C_2 \sum_x |\delta_n^*(x)| M_{j,k,l}(x) f_{\theta^*}(x) \\ &= C_2 \sum_x r_n^*(x) M_{j,k,l}(x) \quad (\text{by CLT}) \\ &\rightarrow C_2 E_g[M_{j,k,l}(X)] < \infty. \quad [\text{by assumption (A6)}] \end{aligned} \quad (27)$$

Thus term (i) is bounded. Now for the second term (ii), we again use (13) to get

$$\begin{aligned}
& \left| \sum_x K'(\delta_n^*(x)) \delta_n^*(x) f_{\theta^*}^{1+\alpha}(x) u_{j\theta^*}(x) u_{k\theta^*}(x) u_{l\theta^*}(x) \right| \\
& \leq C_1 \sum_x |\delta_n^*(x)| M_{j,k,l}(x) f_{\theta^*}(x) \\
& = C_1 \sum_x r_n^*(x) M_{j,k,l}(x) \quad (\text{by CLT}) \\
& \rightarrow C_1 E_g[M_{j,k,l}(X)] < \infty, \quad [\text{by assumption (A6)}]
\end{aligned} \tag{28}$$

so that term (ii) is also bounded. Similarly the terms (iii), (iv) and (v) are bounded as in case of term (ii) and using (13) and assumption (SA6). Next for the term (vi), we will consider the following:

$$|K(\delta)| = \left| \int_0^\delta K'(\delta) d\delta \right| \leq C_1 |\delta|, \tag{29}$$

so that

$$|K(\delta_n^*(x))| \leq C_1 \frac{r_n(x)}{f_{\theta^*}(x)}. \tag{30}$$

Also,

$$\begin{aligned}
\frac{\nabla_{jkl} f_{\theta^*}(x)}{f_{\theta^*}(x)} &= u_{jkl\theta^*}(x) + u_{jk\theta^*}(x) u_{l\theta^*}(x) + u_{jl\theta^*}(x) u_{k\theta^*}(x) \\
&\quad + u_{j\theta^*}(x) u_{kl\theta^*}(x) + u_{j\theta^*}(x) u_{k\theta^*}(x) u_{l\theta^*}(x).
\end{aligned} \tag{31}$$

So,

$$\begin{aligned}
\left| \sum_x K(\delta_n^*(x)) f_{\theta^*}^{1+\alpha}(x) \frac{\nabla_{jkl} f_{\theta^*}(x)}{f_{\theta^*}(x)} \right| &\leq C_1 \sum_x r_n(x) f_{\theta^*}^\alpha(x) \left| \frac{\nabla_{jkl} f_{\theta^*}(x)}{f_{\theta^*}(x)} \right| \\
&= C_1 \sum_x r_n^*(x) M(x) \quad (\text{by CLT}) \\
&\rightarrow C_1 E_g[M(X)] < \infty \\
&\quad [\text{by assumption (SA6)}],
\end{aligned} \tag{32}$$

where  $M(x) = M_{jkl}(x) + M_{jk,l}(x) + M_{jl,k}(x) + M_{j,k,l}(x) + M_{j,k,l}(x)$ . Thus the term (vi) is bounded and also similarly the term (vii) is bounded. Hence, we have  $|S_3| < ba^3$  on the sphere  $Q_a$  with probability tending to one. Combining the three inequalities we get that

$$\max(S_1 + S_2 + S_3) < -ca^2 + (b+p)a^3,$$

which is strictly negative for  $a < \frac{c}{b+p}$ . Thus, for any sufficiently small  $a$ , there exists a sequence of roots  $\theta_n = \theta_n(a)$  to the minimum  $S$ -divergence estimating equation such that  $P(\|\theta_n - \theta^g\|_2 < a)$  converges to one, where  $\|\cdot\|_2$  denotes the  $L_2$ -norm.

It remains to show that we can determine such a sequence independent of  $a$ . For let  $\theta_n^*$  be the root which is closes to  $\theta^g$ . This exists because the limit of a sequence of roots is again a root by the continuity of the  $S$ -divergence. This completes the proof of the consistency part.

**Proof of the asymptotic Normality:** For the Asymptotic normality, we expand

$$\sum_x K(\delta_n(x)) f_{\theta}^{1+\alpha}(x) u_{\theta}(x)$$

in Taylor series about  $\theta = \theta^g$  to get

$$\begin{aligned}
& \sum_x K(\delta_n(x)) f_\theta^{1+\alpha}(x) u_\theta(x) \\
&= \sum_x K(\delta_n^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{\theta^g}(x) \\
&\quad + \sum_k (\theta_k - \theta_k^g) \nabla_k \left( \sum_x K(\delta_n(x)) f_\theta^{1+\alpha}(x) u_\theta(x) \right) \Big|_{\theta=\theta^g} \\
&\quad + \frac{1}{2} \sum_{k,l} (\theta_k - \theta_k^g)(\theta_l - \theta_l^g) \nabla_{kl} \left( \sum_x K(\delta_n(x)) f_\theta^{1+\alpha}(x) u_\theta(x) \right) \Big|_{\theta=\theta'},
\end{aligned} \tag{33}$$

where,  $\theta'$  lies in between  $\theta$  and  $\theta^g$ .

Now, let  $\theta_n$  be the solution of the minimum  $S$ -divergence estimating equation, which can be assumed to be consistent by the previous part. Replace  $\theta$  by  $\theta_n$  in above (33) so that the LHS of the equation becomes zero and hence we get

$$\begin{aligned}
& -\sqrt{n} \sum_x K(\delta_n^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{\theta^g}(x) \\
&= \sqrt{n} \sum_k (\theta_n k - \theta_k^g) \times \left\{ \nabla_k \left( \sum_x K(\delta_n(x)) f_\theta^{1+\alpha}(x) u_\theta(x) \right) \Big|_{\theta=\theta^g} \right. \\
&\quad \left. + \frac{1}{2} \sum_l (\theta_n l - \theta_l^g) \nabla_{kl} \left( \sum_x K(\delta_n(x)) f_\theta^{1+\alpha}(x) u_\theta(x) \right) \Big|_{\theta=\theta'} \right\}.
\end{aligned} \tag{34}$$

Note that, the first term within the bracketed quantity in the RHS of above (34) converges to  $J_g$  with probability tending to one, while the second bracketed term is an  $o_p(1)$  term (as proved in the proof of consistency part). Also, by using the Lemma 4.4, we get that

$$\begin{aligned}
& \sqrt{n} \sum_x K(\delta_n^g(x)) f_{\theta^g}^{1+\alpha}(x) u_{\theta^g}(x) \\
&= \sqrt{n} \sum_x [K(\delta_n^g(x)) - K(\delta_g^g(x))] f_{\theta^g}^{1+\alpha}(x) u_{\theta^g}(x) \\
&= S_{1n}|_{\theta=\theta^g} \xrightarrow{\mathcal{D}} N_p(0, V_g).
\end{aligned} \tag{35}$$

Therefore, by Lehmann (1983, Lemma 4.1), it follows that  $\sqrt{n}(\theta_n - \theta^g)$  has asymptotic distribution as  $N_p(0, J_g^{-1} V_g J_g^{-1})$ .  $\square$

**Corollary 4.6** *When the true distribution  $G$  belongs to the model family, i.e.,  $G = F_\theta$  for some  $\theta \in \Theta$ , then  $\sqrt{n}(\theta_n - \theta)$  has asymptotic distribution as  $N_p(0, J^{-1} V J^{-1})$ , where*

$$J = J_\alpha(f_\theta) = E_{f_\theta}[u_\theta(X) u_\theta(X)^T f_\theta^\alpha(X)] = \int u_\theta(x) u_\theta^T(x) f_\theta^{1+\alpha}(x) dx, \tag{36}$$

$$V = V_\alpha(f_\theta) = V_{f_\theta}[u_\theta(X) f_\theta^\alpha(X)] = \int u_\theta(x) u_\theta^T(x) f_\theta^{1+2\alpha}(x) dx - \xi \xi^T, \tag{37}$$

$$\xi = \xi_\alpha(f_\theta) = E_{f_\theta}[u_\theta(X) f_\theta^\alpha(X)] = \int u_\theta(x) f_\theta^{1+\alpha}(x) dx. \tag{38}$$

Note that, this asymptotic distribution is independent of the parameter  $\lambda$  in the  $S$ -divergence Family.

**Proof:** Note that, under  $G = F_\theta$  for some  $\theta \in \Theta$ , we get  $\delta_g^g(x) = 1 \quad \forall x$  so that  $K(\delta_g^g(x)) = 1^A - 1 = 0$  and  $K'(\delta_g^g(x)) = A1^{(A-1)} = A$ . Thus,

$$\begin{aligned} J_g &= A \int u_\theta(x) u_\theta^T(x) f_\theta^{1+\alpha}(x) dx = AJ, \\ V_g &= A^2 V_g[u_\theta(X) f_\theta^\alpha(X)] = A^2 V, \end{aligned}$$

so that  $J_g^{-1} V_g J_g^{-1} = J^{-1} V J^{-1}$  and the result follows from the above theorem.  $\square$

## 5 Real Data Example

Here we consider a chemical mutagenicity experiment. These data were analyzed previously by Simpson (1987). The details of the experimental protocol are available in Woodruff et al. (1984). In a sex linked recessive lethal test in *Drosophila* (fruit flies), the experimenter exposed groups of male flies to different doses of a chemical to be screened. Each male was then mated with unexposed females. Sampling 100 daughter flies from each male (roughly), the number of daughters carrying a recessive lethal mutation on the  $X$  chromosome was noted. The data set consisted of the observed frequencies of males having 0, 1, 2,  $\dots$  recessive lethal daughters. For our purpose, we consider two specific experimental runs — one on the day 28 and second on day 177. The data of the first run consist of two small outliers with observed frequencies  $d = (23, 3, 1, 1)$  at  $x = (0, 1, 3, 4)$  and that of second run consists of observed frequencies  $d = (23, 7, 3, 1)$  at  $x = (0, 1, 2, 91)$  with a large outlier at 91.

Poisson models are fitted to the data for this experimental runs by estimating the Poisson parameter using minimum  $S$ -divergence estimation for several values of  $\alpha$  and  $\lambda$ . A quick look at the observed frequencies for the experimental run reveals that there is an exceptionally large count – where one male is reported to have produced 91 daughters with the recessive lethal mutation. We estimate the Poisson parameter from this data with the outlying observation and without that outlying observation. The difference in these two estimates gives an indication of the robust behavior (or lack thereof) of different Minimum  $S$ -divergence estimators. Our findings are reported in Table [1] to [4]. We can see from the tables that the estimates differ significantly in the presence of both types of outliers for all  $\alpha$  with  $\lambda > 0$  and for large  $\alpha$  with  $\lambda \geq 0$  as seen in the above simulation study.

## 6 Conclusion

The  $S$ -divergence family generates a large class of divergence measures having several important properties. Thus, the minimum divergence estimators obtained by minimizing these different members of the  $S$ -divergences family also have several interesting properties in terms of their efficiency and robustness. In this present paper, we have proved the asymptotic properties of the minimum  $S$ -divergence estimators under the discrete set-up. Interestingly, we have seen that the asymptotic distributions of the minimum  $S$ -divergence estimators at the model is independent of one defining parameter  $\lambda$ , although their robustness depends on this parameter value. Indeed, considering the minimum  $S$ -divergence estimators as members of a grid constructed based on its defining parameters  $\lambda$  and  $\alpha$ , we can clearly observe a triangular region of non-robust estimators corresponding to the large  $\lambda$  and small  $\alpha$  values and a region of highly robust estimators corresponding to moderate  $\alpha$  and large negative  $\lambda$  values. As a future work, we need to prove all the properties of the minimum  $S$ -divergence estimators under the continuous models. However, under the continuous model, we need to use the kernel smoothing to estimate the true density  $g$  and hence proving the asymptotic properties will inherit all the complications of the kernel estimation like bandwidth selection etc. We will try to solve these issues in our subsequent papers.

Table 1: The estimate of the Poisson parameter for different values of  $\alpha$  and  $\lambda$  for Drosophila data without outlier: First Experimental Run

$\lambda$	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.25$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1$
-1	-	0.08	0.11	0.12	0.12	0.12	0.13	0.13
-0.7	0.09	0.10	0.12	0.12	0.12	0.13	0.13	0.13
-0.5	0.10	0.11	0.12	0.12	0.12	0.13	0.13	0.13
-0.3	0.11	0.12	0.12	0.12	0.12	0.13	0.13	0.13
-0.1	0.11	0.12	0.12	0.12	0.13	0.13	0.13	0.13
0	0.12	0.12	0.12	0.12	0.13	0.13	0.13	0.13
0.5	0.12	0.12	0.12	0.13	0.13	0.13	0.13	0.13
1	0.12	0.12	0.13	0.13	0.09	0.13	0.13	0.13
1.3	0.12	0.12	0.13	0.13	0.13	0.13	0.13	0.13
1.5	0.12	0.12	0.13	0.13	0.13	0.13	0.13	0.13
2	0.12	0.13	0.13	0.13	0.13	0.13	0.13	0.13

Table 2: The estimate of the Poisson parameter for different values of  $\alpha$  and  $\lambda$  for Drosophila data with outlier: First Experimental Run

$\lambda$	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.25$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1$
-1	-	0.08	0.11	0.13	0.14	0.14	0.15	0.16
-0.7	0.10	0.11	0.13	0.14	0.14	0.15	0.16	0.16
-0.5	0.13	0.13	0.13	0.14	0.14	0.15	0.16	0.16
-0.3	0.18	0.15	0.14	0.14	0.14	0.15	0.16	0.16
-0.1	0.29	0.22	0.16	0.15	0.15	0.15	0.16	0.16
0	0.36	0.26	0.18	0.15	0.15	0.15	0.16	0.16
0.5	0.59	0.49	0.34	0.21	0.17	0.16	0.16	0.16
1	0.70	0.63	0.49	0.32	0.18	0.17	0.16	0.16
1.3	0.75	0.68	0.55	0.39	0.28	0.19	0.16	0.16
1.5	0.77	0.71	0.59	0.44	0.32	0.25	0.16	0.16
2	0.81	0.76	0.66	0.52	0.40	0.27	0.16	0.16

Table 3: The estimate of the Poisson parameter for different values of  $\alpha$  and  $\lambda$  for Drosophila data without outlier: Second Experimental Run

$\lambda$	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.25$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1$
-1	-	0.29	0.35	0.36	0.36	0.35	0.35	0.35
-0.7	0.34	0.35	0.36	0.36	0.36	0.36	0.35	0.35
-0.5	0.36	0.37	0.37	0.36	0.36	0.36	0.35	0.35
-0.3	0.38	0.38	0.37	0.37	0.36	0.36	0.35	0.35
-0.1	0.39	0.39	0.38	0.37	0.37	0.36	0.35	0.35
0	0.39	0.39	0.38	0.37	0.37	0.36	0.35	0.35
0.5	0.41	0.40	0.39	0.38	0.37	0.36	0.35	0.35
1	0.42	0.42	0.40	0.39	0.32	0.37	0.36	0.35
1.3	0.43	0.42	0.41	0.39	0.38	0.37	0.36	0.35
1.5	0.43	0.42	0.41	0.39	0.38	0.37	0.36	0.35
2	0.44	0.43	0.42	0.40	0.39	0.37	0.36	0.35

Table 4: The estimate of the Poisson parameter for different values of  $\alpha$  and  $\lambda$  for Drosophila data with outlier: Second Experimental Run

$\lambda$	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.25$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1$
-1	—	0.30	0.35	0.36	0.36	0.36	0.36	0.36
-0.7	0.34	0.36	0.37	0.37	0.37	0.37	0.36	0.36
-0.5	0.36	0.37	0.37	0.37	0.37	0.37	0.37	0.36
-0.3	0.38	0.38	0.38	0.37	0.37	0.37	0.37	0.36
-0.1	0.39	0.39	0.38	0.38	0.37	0.37	0.37	0.36
0	3.03	0.39	0.39	0.38	0.37	0.37	0.37	0.36
0.5	31.31	30.28	25.12	0.39	0.38	0.37	0.37	0.36
1	32.20	31.84	30.79	27.08	0.99	0.38	0.37	0.36
1.3	32.40	32.15	31.48	29.71	24.93	0.38	0.37	0.36
1.5	32.50	32.29	31.76	30.48	27.78	22.54	0.37	0.36
2	33.22	32.50	32.15	31.43	30.28	26.24	0.37	0.36

## References

- [1] Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika*, **85**, 549–559.
- [2] Basu, A., Shioya, H. and Park, C. (2011). *Statistical Inference: The Minimum Distance Approach*. Chapman & Hall/CRC, Boca Raton, Florida.
- [3] Chung, K.-L. (1974). *A Course in Probability Theory*. New York, USA: Academic Press.
- [4] Cressie, N. and T. R. C. Read (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society B*, **46**, 440–464.
- [5] Csizsár, I. (1963). Eine informations theoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publ. Math. Inst. Hungar. Acad. Sci.* **3**, 85–107.
- [6] Ghosh, A., Harris, I. R., Maji, A., Basu, A., Pardo, L. (2013). A Generalized Divergence for Statistical Inference. *Technical Report, BIRU/2013/3*, Bayesian and Interdisciplinary Research Unit, Indian Statistical Institute, Kolkata, India.
- [7] Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.
- [8] Lindsay, B. G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Annals of Statistics*, **22**, 1081–1114.
- [9] Simpson, D. G. (1987). Minimum Hellinger distance estimation for the analysis of count data. *Journal of the American Statistical Association*, **82**, 802–807.
- [10] Woodruff, R. C., J. M. Mason, R. Valencia, and A. Zimmering (1984). Chemical mutagenesis testing in drosophila I: Comparison of positive and negative control data for sex-linked recessive lethal mutations and reciprocal translocations in three laboratories. *Environmental Mutagenesis*, **6**, 189–202.